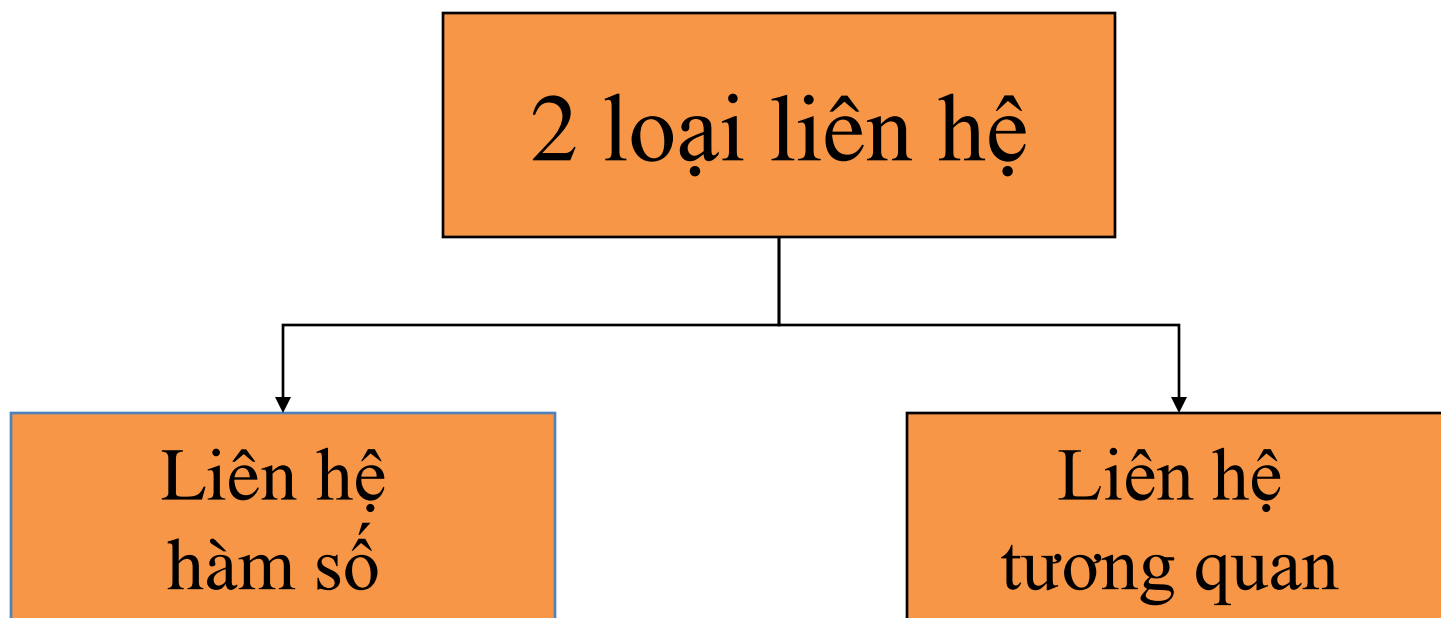


CHƯƠNG 4

HỒI QUI VÀ TƯƠNG QUAN

I - Mối liên hệ giữa các hiện tượng và nhiệm vụ của phương pháp hồi qui và tương quan.

1 - Mối liên hệ giữa các hiện tượng



- Liên hệ hàm số

- + Mỗi liên hệ hoàn toàn chặt chẽ và được biểu hiện dưới dạng một hàm số $y = f(x)$ (sự biến đổi của x hoàn toàn quyết định sự thay đổi của y).
- + Không chỉ thấy được trên toàn bộ tổng thể mà còn thấy được trên từng đơn vị riêng biệt.
- + VD : $S = v.t$

- Liên hệ tương quan

- + Mỗi liên hệ không hoàn toàn chặt chẽ giữa các hiện tượng nghiên cứu.
- + Thường không biểu hiện rõ trên từng đơn vị cá biệt, do đó cần nghiên cứu hiện tượng số lớn.
- + Phương pháp dùng nghiên cứu mỗi liên hệ tương quan là **phương pháp hồi qui và tương quan**.

2- Nhiệm vụ của phương pháp hồi qui và tương quan

* Nhiệm vụ tổng quát:

- Là phương pháp toán học được vận dụng trong thống kê để biểu hiện và phân tích mối liên hệ tương quan giữa các hiện tượng kinh tế xã hội.

Nhiệm vụ cụ thể (nội dung của pp hồi qui và tương quan):

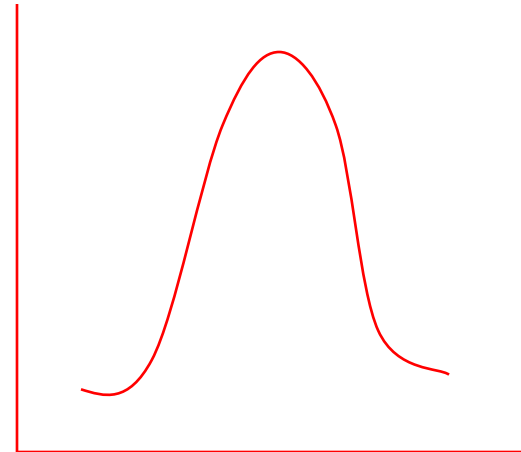
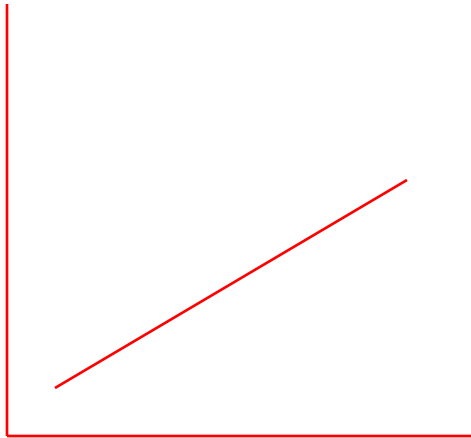
a/ Xác định phương trình hồi qui: 4 bước

B1 : Dựa vào phân tích lý luận để giải thích sự tồn tại thực tế và bản chất của mối liên hệ:

- + Các tiêu thức nghiên cứu có liên hệ không
- + Xác định tiêu thức nguyên nhân, tiêu thức kết quả

B2 : Xác định hình thức, tính chất của mối liên hệ.

- Hình thức : thuận hay nghịch
- Tính chất : Tuyến tính hay phi tuyến tính



B3 : Lập phương trình hồi qui biểu diễn mối liên hệ.

B4 : Tính toán các tham số, giải thích ý nghĩa các tham số.

b/ Đánh giá trình độ chặt chẽ của mối liên hệ

- Hệ số tương quan
- Tỷ số tương quan.

II – Liên hệ tương quan tuyến tính giữa 2 tiêu thức

VD : Có số liệu sau (thu thập từ 10 SV được chọn một cách ngẫu nhiên):

Xác định mối liên hệ giữa số buổi vắng mặt và điểm bq bằng phương pháp hồi qui và tương quan

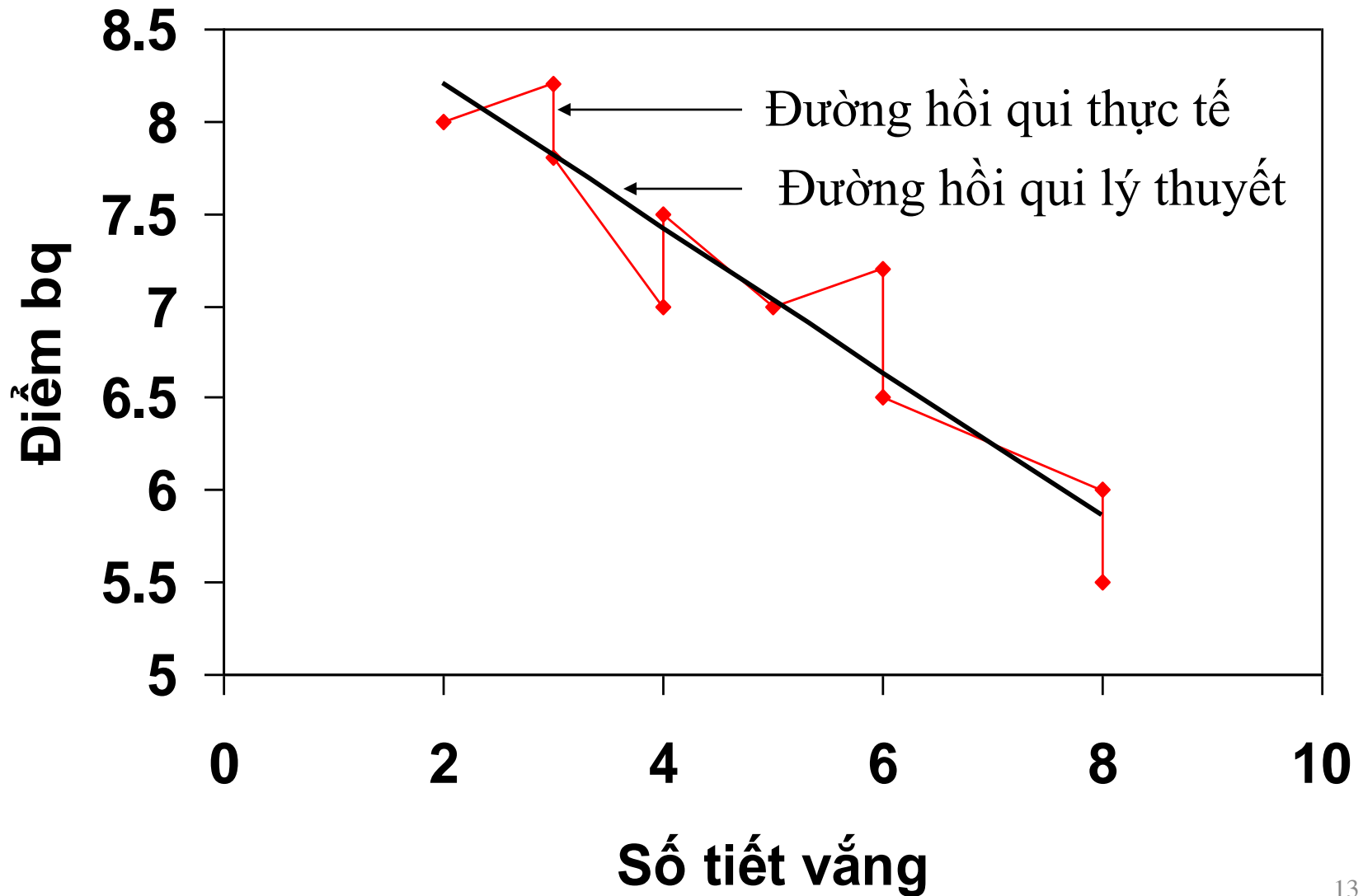
STT	Số tiết vắng mặt	Điểm bình quân
1	3	8,2
2	4	7,0
3	5	7,0
4	6	7,2
5	8	5,5
6	3	7,8
7	4	7,5
8	6	6,5
9	2	8,0
10	8	6,0

1 – Xác định phương trình hồi qui

- Sắp xếp thứ tự và vẽ đồ thị:

Số tiết vắng mặt (x)	Điểm bình quân (y)
2	8,0
3	8,2
3	7,8
4	7,0
4	7,5
5	7,0
6	7,2
6	6,5
8	6
8	5,5

Vẽ đồ thị



Phương trình hồi qui:

$$y_x = a + bx$$

Trong đó :

x : Trị số của tiêu thức nguyên nhân

y_x : Trị số điều chỉnh của tiêu thức kết quả y theo quan hệ phụ thuộc với x

a, b : Các tham số

a : tham số tự do nói lên ảnh hưởng của các nguyên nhân khác ngoài x đối với y

b : Hệ số hồi qui, phản ánh độ dốc của đường hồi qui và nói lên ảnh hưởng của x đối với y , cụ thể mỗi khi x tăng 1 đơn vị thì y tăng bình quân b đơn vị.

Xác định a,b dựa vào phương pháp bình phương nhỏ nhất.

* Phương pháp bình phương nhỏ nhất:

Tối thiểu hoá tổng bình phương các độ lệch giữa giá trị thực tế và giá trị điều chỉnh của biến phụ thuộc y.

C1 : Tính a,b từ hệ phương trình

$$\begin{cases} \sum y = na + b \sum x \\ \sum xy = a \sum x + b \sum x^2 \end{cases}$$

C2 : Tính a , b theo công thức :

$$b = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x^2}$$

$$a = \bar{y} - b\bar{x}$$

Tính lại cho VD : $\sigma_x^2 = 3,89$

$$b = -0,3915$$

$$a = 8,988$$

2 - Hệ số tương quan (r)

- Ý nghĩa : Đánh giá trình độ chặt chẽ của mối liên hệ tương quan tuyến tính.
- CT :

$$r = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \cdot \sigma_y}$$

$$r = b \cdot \frac{\sigma_x}{\sigma_y}$$

- Tác dụng của r
- + Xác định cường độ của mối liên hệ
- + Xác định phương hướng của mối liên hệ.
 - $r > 0$: liên hệ tương quan thuận
 - $r < 0$: liên hệ tương quan nghịch
- + Dùng nhiều trong phân tích và dự đoán TK

- Tính chất của r : $-1 \leq r \leq 1$

+ $r = 1$: Giữa x và y có mối liên hệ hàm số

+ $r = 0$: Giữa x và y không có mối liên hệ tương quan tuyến tính.

+ r càng tiến gần tới 1 : Mối liên hệ giữa x và y càng chặt chẽ.

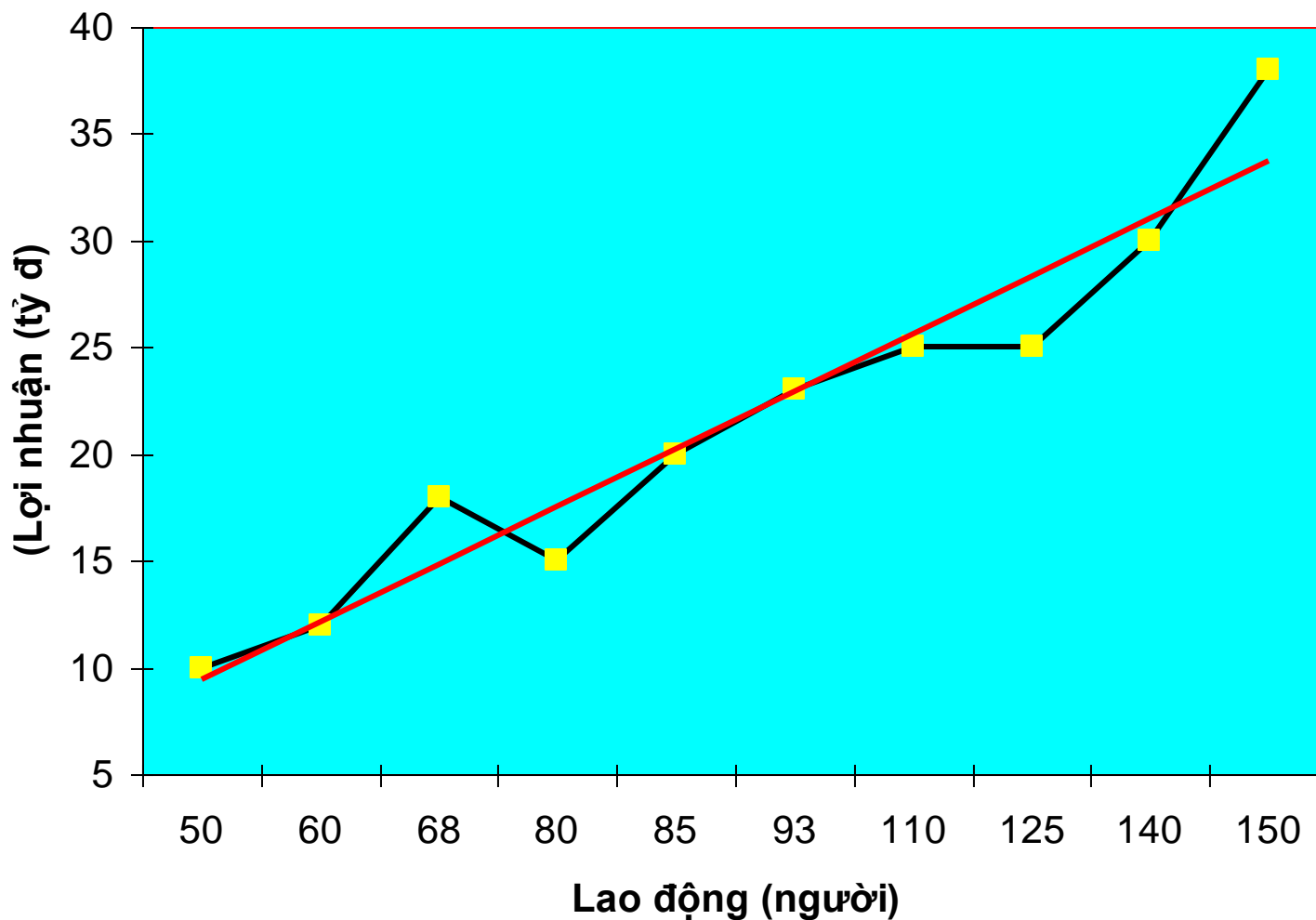
Tính r cho VD , kq $r = -0,93$

KL ?

Bài tập

Số lao động (người)	50	60	68	80	85	93	110	125	140	150
Lợi nhuận (tỷ đ)	10	12	18	15	20	23	25	25	30	38

➤ Biểu diễn mối liên hệ giữa 2 tiêu thức



III – Liên hệ tương quan phi tuyến tính giữa 2 tiêu thức số lượng

1 – Xác định phương trình hồi qui

a/ Phương trình parabol bậc 2

$$y_x = a + bx + cx^2$$

Hệ phương trình để xđ a,b,c:

$$\begin{cases} \sum y = na + b \sum x + c \sum x^2 \\ \sum xy = a \sum x + b \sum x^2 + c \sum x^3 \\ \sum x^2 y = a \sum x^2 + b \sum x^3 + c \sum x^4 \end{cases}$$

b/ Phương trình hypebol:

$$y_x = a + b \cdot \frac{1}{x}$$

Xác định a, b dựa trên hệ phương trình:

$$\sum y = na + b \sum \frac{1}{x}$$

$$\sum \frac{y}{x} = a \sum \frac{1}{x} + b \sum \frac{1}{x^2}$$

2 – Tỷ số tương quan (η) - êta

Ý nghĩa : Đánh giá trình độ chặt chẽ của mối liên hệ tương quan phi tuyến tính.

CT :

$$\eta = \sqrt{\frac{\sigma_{y_x}^2}{\sigma_y^2}} = \sqrt{1 - \frac{\sigma_{y(x)}^2}{\sigma_y^2}} = \sqrt{1 - \frac{\sum (y - y_x)^2}{\sum (y - \bar{y})^2}}$$

Trong đó :

$$\sigma_y^2 = \frac{\sum (y - \bar{y})^2}{n} = \frac{\sum y^2}{n} - (\bar{y})^2$$

$$\sigma_{y_x}^2 = \frac{\sum (y_x - \bar{y})^2}{n}$$

$$\sigma_{y(x)}^2 = \frac{\sum (y - y_x)^2}{n}$$

Quan hệ giữa 3 phương sai ?

Tính chất của η :

Tỷ số tương quan có giá trị nằm trong khoảng $[0 ; 1]$

$$0 \leq \eta \leq 1$$

Nếu $\eta = 0$: x, y không có mối liên hệ tương quan

Nếu $\eta = 1$: x, y có liên hệ hàm số

Nếu η càng gần 1 thì liên hệ tương quan càng chặt chẽ

IV – Tương quan tuyến tính đa biến

a/ Phương trình hồi qui :

$$y_{x_1 x_2 \dots x_n} = a_0 + a_1 x_1 + a_2 x_2 + \dots + a_n x_n$$

Xđ $a_0, a_1, a_2, \dots, a_n$ dựa trên hệ phương trình:

$$\left\{ \begin{array}{l} \sum y = ma_0 + a_1 \sum x_1 + a_2 \sum x_2 + \dots + a_n \sum x_n \\ \sum x_1 y = a_0 \sum x_1 + a_1 \sum x_1^2 + a_2 \sum x_1 x_2 + \dots + a_n \sum x_1 x_n \\ \sum x_2 y = a_0 \sum x_2 + a_1 \sum x_1 x_2 + a_2 \sum x_2^2 + \dots + a_n \sum x_2 x_n \\ \dots\dots\dots \\ \sum x_n y = a_0 \sum x_n + a_1 \sum x_1 x_n + a_2 \sum x_2 x_n + \dots + a_n \sum x_n^2 \end{array} \right.$$

b/ Hệ số tương quan bội :

Ý nghĩa : Đánh giá trình độ chặt chẽ của mối liên hệ tương quan tuyến tính đa biến.

Công thức :

$$\begin{aligned} R_{yx_1x_2 \dots x_n} &= \sqrt{\frac{\sigma_{yx_1x_2 \dots x_n}^2}{\sigma_y^2}} = \sqrt{1 - \frac{\sigma_{y(x_1x_2 \dots x_n)}^2}{\sigma_y^2}} \\ &= \sqrt{1 - \frac{\sum (y - y_{x_1x_2 \dots x_n})^2}{\sum (y - \bar{y})^2}} \end{aligned}$$

- Tính chất của hệ số tương quan bội
Có giá trị nằm trong khoảng $[0 ; 1]$

$$0 \leq R_{y_{x_1 x_2 \dots x_n}} \leq 1$$

Nếu $R = 0$: Giữa y và các x_1, x_2, \dots, x_n không có liên hệ tuyến tính.

Nếu $R = 1$: Giữa y và các x_1, x_2, \dots, x_n có liên hệ hàm số.

R càng gần 1, mối liên hệ giữa y và các x_1, x_2, \dots, x_n càng chặt chẽ.

- Tham số tương quan chuẩn hoá (β)

Ý nghĩa : Đánh giá mức độ ảnh hưởng của từng biến độc lập x_i tới biến phụ thuộc y .

Công thức :

$$\beta_k = a_k \cdot \frac{\sigma_{x_k}}{\sigma_y}$$